MFPaQ documentation

1) <u>Snapshot</u>

MFPaQ (Mascot File Parsing and Quantification) is a new software developed on the IPBS (Institut de Pharmacologie et de Biologie Structurale, Toulouse, France) proteomics platform and dedicated to parse, validate, and quantify proteomic data. It allows fast and user-friendly verification of Mascot result files, as well as data quantification from an experiment performed by isotopic labeling using either ICAT or SILAC methods.

This new tool provides a convenient interface to retrieve Mascot protein lists, sort them according to Mascot scoring or to user-defined criteria based on the number, the score and the rank of identified peptides, and to validate the results. The software extracts quantitative data from raw files obtained by nanoLC-MS/MS, calculates peptide ratios, and generates a non-redundant list of proteins identified in a multi-search experiment with their calculated averaged and normalized ratio.

It is based on three modules, the Mascot File Parser (MFP) Module, the quantification module and a third module designed for differential analysis, in which validated protein lists are compared. The input of the MFP module is a list of mascot .dat files and the input of the quantification module is a list of .wiff files generated by Analyst QS on a Qstar instrument (coming soon: version of the quantification module compatible with other mass spectrometers).

The installation package is available in two different versions:
- the first one contain only the program files (quick download),
- the other one comes with the example files presented in the Technology Paper T6:00069-MCP, entitled "MFPaQ, a new software to parse, validate, and quantify proteomic data generated by ICAT and SILAC mass spectrometric analyses: application to the proteomic study of membrane proteins from primary human endothelial cells" (this version will allow you to view all the results of this study in an "example" profile already created)

The next section provides details about the downloading and installation processes.

2) <u>Download and installation instructions</u>

2.1 Download
The application can be downloaded from http://sourceforge.net/projects/mfpaq

2.2 Prerequisites

MFPaQ is a web-based application that runs on a server with Windows XP Pro edition and Windows 2003 Server. Its installation needs some other softwares/packages to be installed:
- IIS (Internet Information Services): this is a component of Windows that can be installed from the "add/remove programs" (configuration panel). Select "add/remove windows components", check "IIS", click "next" and validate (you may have to insert the Windows CD-ROM).
- Mascot server 2.1 or above
- Analyst QS 1.1 or 2.0
- Perl 5.8 or above (modules: XML-Simple, GD, Spreadsheet::WriteExcel),
- Microsoft.Net Framework 1.1 or above

Remark: Analyst QS is needed only if you wish to use the quantification module. Mascot is needed if you want to use the MFP module to parse .dat files generated by this search engine. However, in order to get only a demo of the software, it is possible to download the second version of the program with the example files and to install it on a computer without these two softwares, provided that IIS, Perl, and Microsoft .Net Framework are installed.

2.3 Installation

2.3.1 Execution of the automated-installer

The downloaded file is an executable that will help you through the installation process. When it is launched, it detects the presence of the needed softwares, then copy the application files in the appropriate directory. It will show a message if a needed component is absent. If Mascot is present on the computer, the application will be installed in the "data" sub-directory of Mascot. Otherwise, the application will be installed in the web-server "root" directory.

2.3.2 IIS Web-Server configuration

The configuration panel is automatically displayed at the end of the installation process. If not, this panel is available through:
configuration panel => administration tools => Internet Services (IIS).

In order to execute Perl scripts on the server, you need to configure the extension ".pl". To do this, right click on the "default website" of the IIS configuration panel and click on "properties". In the base directory tab, click on configuration then click on add button.
Fill the fields like this:
- executable: C:\Perl\bin\perl.exe "%s" %s
- extension: .pl
Notes:
- it is considered that Perl is installed on C:\.
- if you have an ActiveState Perl distribution, the ".pl" extension may have been already configured.

Important: in the same tab, set the local path to the mascot data directory.
The "mfpaq" directory is now a part of the website folder tree.

You have to set the security configuration of the application. Here is a table containing the access rights corresponding to each directory:

Folders	Access	Autorisations
mfpaq	Read	none
mfpaq/cgi-bin	Read	scripts and executables
mfpaq/img	Read	none
mfpaq/Scripts	Read	scripts only
mfpaq/styles	Read	none
mfpaq/xml	Read/Write	none

To set them right click on each application folder then click on properties in the directory tab.

2.3.3 Windows configuration

Decimal symbol as to be the point caracter (see windows regional settings).
Important: the windows group "User" needs to have a total control on the mfpaq/xml directory.

3) Getting started

The user interface is accessible via a web browser: Microsoft Internet Explorer and Mozilla Firefox are currently compatible with the application.

The application is available using one of theses web-browsers and typing in the address bar the name of the mascot web-server followed by '/mfpaq' (EX: **http://myserver/mfpaq** ; this assumes that Mascot is available at **http://myserver/mascot**).

The MFPaQ home page will be displayed and you are ready to use the application. To get started, follow theses steps:

1. Create a new user profile in the menu "Profiles".

2. In order to start a session, go back to the "Home" page, select your profile name in the list of profiles and click on the "Open a session" button.

3. In order to parse Mascot dat files, you need first to set up the parsing criteria that you wish to use. Click on "Configure a profile" in the "Profiles" menu and configure your profile by setting the filters based on Mascot criteria.
The MFP module will extract protein entries from Mascot files and rank them according to either the Mascot "Standard scoring" or "Mudpit scoring". Choose which type of scoring you wish to apply.
In order to facilitate manual validation, the software applies to the proteins of the list a two-colour code related to the filtering rules defined by the user under its configuration profile. Proteins that passed the "validation criteria" are displayed in green. They can be considered as confident hits that do not need further verification and will automatically be checked in the validation window. Proteins that meet the "exclusion criteria" are discarded and are not displayed in the list. All other proteins, which are considered as ambiguous identifications, appear in red and can be manually verified by the user.
The filtering rules used for the classification of a protein in green and red are based either on the protein score defined in Mascot (Click on "Filter according to protein scores"), or on multiple criteria related to the peptide matches assigned to this protein (Click on "Filter according to peptide scores"). In the first case, the software basically displays in green colour the "significant hits" list given in the Mascot (proteins with total scores higher than the significance threshold, which depends on the database size and is calculated by default with the probability for a match to occur at random with a probability of less than 5%). In the other case, if you choose to apply criteria based on the number, the rank, and the score of the peptide matches assigned to a protein, the proteins displayed in MFPaQ will still be ranked according to Mascot scoring, but only proteins matching this criteria will appear in green and will be automatically validated. Proteins that do not fulfill these criteria, while being in the significant list of Mascot, will appear in red and will have to be verified manually.
Once you have chosen which criteria you want to apply, click the "save" button.

4. You can now extract and validate your Mascot results with the "Mascot File Parser" module.

Menu description:

◆ Mascot File Parser:

➥ **"Create and validate an experiment" to parse Mascot .dat files.**
The MFP module offers the possibility to create an "experiment" corresponding to the extraction of one or several Mascot results files (.dat files). In a shotgun analysis, if the protein mixture is fractionated first (e.g. in a series of 1D gel slices) and each protein fraction is digested, then peptides from each fraction will be analyzed by nanoLC-MS/MS. In that case, several Mascot database searches will be performed with the different peaklists obtained from the nanoLC-MS/MS runs, and several Mascot dat result files will be generated . The software extracts in batch mode the data contained in a series of Mascot .dat files specified by the user under an "experiment", and displays a table with links to a validation window for each of these searches

➥ **"Open an experiment" to visualize your Mascot results filtered using your criteria** previously defined in the menu "Configure a profile". You can modify manually and save at any time the validation of the proteins of each result file, by checking the corresponding boxes. A result file that has not been manually validated and saved (only automatic validation has been performed) is displayed in red in the table. If proteins were selected or de-selected by the user in a result data files, and the new list has been saved, then it appears in green in the table. Validated proteins, including all associated peptide information, are saved in XML files and can be exported into Excel. It is also possible to generate an exclusion lists for further nanoLCMS/MS experiments.

➥ **"Generate a protein list"** to edit a list of all previously validated proteins in your experiment. the MFP module can generate a unique, non-redundant list of proteins from all the validated result files of a multi-search experiment. This is particularly useful when protein fractionation is performed, as the same protein can be identified several times in adjacent gel slices. The software compares proteins or protein groups (composed of all the proteins matching the same set of peptides) and creates clusters from protein groups found in different gel slices if they have one common member. This feature allows the editing of a global list of unique proteins (or clusters) representing the entire sample analyzed in the experiment.

➥ **"Open a list"** to access to a protein list already generated.

◆ Quantification:

 Before using this menu you have to run an external module called "Extract Daemon" to retrieve in the MS Survey scan the intensity of the peptides from an experiment previously created and validated with the Mascot File Parser module. In the "Extract Daemon", you will associate each Mascot .dat result file with the corresponding raw MS file, and set up the kind of isotopic labeling you are using. Once this has been done, go back to the "Quantification" menu of MFPaQ.

➡ **"Process a quantification analysis"** to perform the quantitative computation of the selected experiment. The ratios of all validated peptide matches are averaged for each protein in a gel slice, and a coefficient of variation is calculated for the ratio of the proteins which have been quantified with several peptide matches. The software allows the verification of the calculated ratios and the manual de-selection of some peptide pairs or some MS scans in case of aberrant ratio calculation (co-elution with other peptides, weak signal…). Direct links are provided for each protein ratio towards a "Quanti-Viewer" window, showing all data used for quantification of an individual protein. These include the list of isotopically labeled peptide pairs identified for this protein, with peptide score, mass, and elution time, the list of MS scans used to extract peptide intensities, and the corresponding MS spectra of the peptide pairs

➡ **"Open an analysis"** to access to an experiment already analyzed and quantified.

➡ **"Create a quantification report"** to generate a detailed report containing the quantification ratios of each protein in the experiment. In case of protein fractionation, when a protein is identified and quantified several times in consecutive fractions (eg 1D gel slices), a final protein ratio is computed by averaging the different ratios found for this protein in the different fractions, and a global coefficient of variation is calculated. Proteins or protein groups identified and quantified in different fractions are clusterized to generate a final non-redundant list of protein groups, with their normalized protein ratio and the associated global coefficient of variation.

➡ **"Open a quantification report"** to visualize a report.

◆ Differential Analysis:

➡ **"Open a list"** to visualize the lists of validated proteins created in the MFP menu.

➡ **"Compare two lists"** to compare the lists of proteins from two experiments A and B and to generate three new lists:
- one containing the specific proteins from experiment A,
- one containing the specific proteins from experiment B,
- one containing shared proteins from experiments A and B.

➡ **"Merge two lists"** (not yet implemented).

MFPaQ offers an export option to report software results in Microsoft Excel format.

4) <u>Browsing the example files</u>

1) Open the program

2) An "Example" profile has already been generated. To get a demo of the software and view the example datafiles, go in the "Home" page, choose "Example" in the "List of Profiles" menu and open the "Example" session.

3) To view the profile configuration click on "Configure a profile" in the "Profiles" menu. The parameters that were used to parse Mascot .dat files are displayed here.

4) In the "Mascot File Parser" menu, click on "Open experiment" and display the list of experiments. The three experiments described in the manuscript will appear.

5) By clicking on an experiment link (number or name), you reach a table where the Mascot .dat files of the experiment are detailed.

6) To see the parsing of a single .dat file, click on the link, a summary table of proteins found in this mascot result is displayed. You can expand the table with the menu "Show peptides" that display peptide of each protein, and with the menu "Show same sets", that allows to view all the proteins groups (a protein group correspond to all the proteins matching the same set of peptides). When a protein has been automatically validated, it is displayed in green in the table. By default, the check box located on the right of a protein entry is automatically selected for green protein, but manual selection or de-selection of proteins is possible by clicking these boxes, and then saving the manually validated list.

7) To generate concatenated protein lists, go back to the "Mascot file parser" menu and click on "generate a protein list". Choose an experiment by clicking on one experiment name. The concatenated protein list is generated. To open it, click on the link "display the list".

8) "Quantification module": for this set of example, data have been already processed by the "Extract Daemon", an external tool that retrieves the intensity of previously validated peptides in the MS Survey data.
By clicking on the "Quantification" menu, and then on "Open a quantification analysis", you can display the two quantification analyses of the manuscript. The link on each experiment name allows to view the detail of an experiment (list of all mascot dat files, quantified with the corresponding wiff file). If you click on a sample name (in the column "Mascot dat file" or "Title") you reach the quantification window for this particular dat file. Click on the accession number link to display the "Quanti quick-viewer" and see how quantification is performed for each protein.
To generate a global quantification report for an experiment, click on "Create a quantification report" in the "Quantification" menu. Proteins or protein groups identified and quantified in different fractions are clusterized to generate a final non-redundant list of protein groups, with their normalized protein ratio and the associated global coefficient of variation.